



Smart Computer Aided Translation Environment

IWT-SBO 130041

<http://www.ccl.kuleuven.be/scate>

List of consortium partners

University of Leuven (CCL - Centre for Computational Linguistics + ESAT/PSI - Centre for Processing of Speech and Images + LIIR - Language Intelligence & Information Retrieval + Research Unit Linguistics @ Thomas More Antwerp)

University of Ghent (LT3 Language and Translation Technology Team)

Hasselt University (Expertise Centre for Digital Media)

In the SCATE project we aim at improving the translators' efficiency. Commercial translation tools are faced with ever higher productivity requirements imposed by the globalisation of business activities and the increasing information flow.

The SCATE project intends to improve translators' efficiency along the following axes:

- **Exploitation of already translated data** – We will exploit data more exhaustively through the use of syntactic models for fuzzy matching, and detect syntactically similar constructions in the translation memory. We will investigate complex types of translation grammar induction and tree alignment that allow to *transduce* source syntax trees into target trees (i.e. accepting one tree and producing another). We will investigate how to seamlessly integrate MT into a translation memory, by automatically resolving the syntactic fuzziness of the match through MT techniques.
- **Translation evaluation** – We will automatically judge whether MT output is worth post-editing, or whether the suggested translation can be applied to resolve the fuzzy match in the translation memory. We will build an annotated data set and a taxonomy of typical translation errors and combine this with loggings and analysis of human-machine interaction during post-editing, which targets improvements in automatic confidence estimation of machine translation output.
- **Terminology extraction** – We will automatically extract terminology from comparable corpora in order to speed up the translation process and make translations more consistent. Therefore we will study translator's methods in acquiring domain terminology. We will also research methods to determine which texts in different languages contain comparable information, and we will improve current methods of terminology extraction from comparable corpora through techniques such as cross-lingual topic modelling.
- **Speech recognition** – We will integrate the language model of the MT engine with the language model of the speech recogniser. We will study the adaptation of the recogniser as an input method for the post-editor, and investigate the improvement of speech transcription for translation purposes. Furthermore we will study how to perform automatic domain-adaptation for speech recognition, in order to automatically adapt the language models of the recogniser to the domain.
- **Workflows and personalised user interfaces** – We aim at a higher comfort and productivity for the translators, by analysing and modelling current translation systems and translator's workflows and practices, investigating new visualisations of translation features, and developing and testing new interfaces for translation work.